63-I-2

MATHEMATICS RESEARCH CENTER, UNITED STATES ARMY
THE UNIVERSITY OF WISCONSIN

Contract No.: DA-11-022-ORD-2059

THE HEURISTICS OF DATA TRANSFORMATION

H. E. Reinhardt

MRC Technical Summary Report #355
February 1963

Madison, Wisconsin

ABSTRACT

Suppose $\{X_n\}$ is a sequence of random variables (r.v.s) with means μ_n and variances $f_n(\mu_n)$. Attempts to find r.v.s $T_n = T_n(X_n)$ with approximately constant variance have focussed on r.v.s of the form

$$(1) \qquad T_n(X_n) = \int_{\mu_n}^{X_n} \frac{1}{\sqrt{f_n(t)}}\,dt \ .$$

Curtiss (Annals of Math. Stat., vol. 14, 107-122) proved a fundamental theorem which gave a sound theoretical basis for transformations of the form (1). This note gives several generalizations and applications of Curtiss's theorem. Typical is the following:

Suppose the sequence of r.v.s $\{X_n - \mu_n\}$ converge in probability to 0, the sequence of real numbers $\{\mu_n\}$ converges to a finite limit μ and the distribution functions of $\{(X_n - \mu_n)\alpha_n\}$ (α_n a positive constant) converges to a d.f. $F(\omega)$. If the real-valued function of a real variable $\xi(x)$ has a continuous derivative $\xi'(x)$ which does not vanish at $x = \mu$, then the d.f's of the sequence

$$\left\{ \frac{\xi(X_n) - \xi(\mu_n)}{\xi'(\mu)}\ \alpha_n \right\}$$

converge to $F(\omega)$.

Standard theorems of real function theory and standard techniques of probability theory are employed.

THE HEURISTICS OF DATA TRANSFORMATION

H. E. Reinhardt[1]

1. Introduction: Suppose $\{X_n\}$ is a sequence of random variables (r.v.s) with means μ_n and variances $f_n(\mu_n)$. Attempts to find r.v.s $T_n = T_n(X_n)$ with approximately constant variance have focussed on r.v.s of the form

$$T_n(X_n) = \int_{\mu_n}^{X_n} \frac{1}{\sqrt{f_n(t)}}\,dt \quad . \tag{1}$$

The heuristic argument usually advanced for such a transformation involves approximating $T_n(x)$ by the linear term of its Taylor series expansion in a neighborhood of μ_n . Of course, heuristic arguments are a matter of taste, but many people have pointed out difficulties connected with this one. An early reference is [3], a recent one is [5], p. 72.

In 1943, Curtiss [3] proved a fundamental theorem which gave a sound theoretical basis to transformations of the form (1).

This note gives an alternate heuristic argument which can be made rigorous. In its rigorous form it is a generalization of Curtiss's theorem and implies many of the standard asymptotic theorems. Standard techniques of real function theory and probability theory are used.

---

[1] Mathematics Research Center, U. S. Army and Montana State University.

[2] Sponsored by the Mathematics Research Center, United States Army, Madison, Wisconsin under Contract No.: DA-11-022-ORD-2059.

2. <u>An Heuristic Argument</u>. We follow Curtiss in his formulation. We consider a sequence of r.v.s $\{X_n\}$, a sequence of real numbers $\{\mu_n\}$, and a sequence of real-valued integrable functions $\varphi_n(x)$ defined on the real numbers, such that the sequence of r.v.s $Y_n = (X_n - \mu_n)\,\varphi_n(\mu_n)$ have distribution functions (d.f.s) which converge to a d.f. $F(\omega)$ at continuity points of F . (Let Y be a r.v. with d.f.F. We shall follow Parzen [7], p. 424, saying that the law of Y_n converges to the law of Y and writing $\mathcal{L}(Y_n) \to \mathcal{L}(Y)$.)

We consider the sequence of r.v.s $T_n(X_n)$. If φ_n is continuous, then, according to the mean value theorem for integrals, $T_n = (X_n - \mu_n)\,\varphi_n(\zeta)$ for suitable ζ . Now if φ_n is a "slowly-changing" function T_n will be approximately $(X_n - \mu_n)\,\varphi_n(\mu_n)$ so that we should have $\mathcal{L}(T_n) \to \mathcal{L}(Y)$.

(Arley and Buch [1], p. 79], in considering the problem of data transformation have applied the adjective "slowly-varying" to the function $T_n(x)$, interpreting this as a condition relating the first two derivatives of T_n, which they assume exist. In making the above argument rigorous, we shall use "slowly-varying" in the precise sense of Curtiss as stated in Theorem 1, below.)

3. <u>The Heuristic Argument Made Rigorous</u>. We use this line of argument to prove Theorem 1. We consider throughout this paper a sequence of random variables $\{X_n\}$, a sequence of real numbers $\{\mu_n\}$, and a sequence of real-valued functions of a real variable $\{\varphi_n(x)\}$, Lebesque integrable with respect to x for each finite interval and each n. We shall be concerned with the following conditions.

Condition A: The Laws of the r.v.s. $Y_n = (X_n - \mu_n)\,\varphi_n(\mu_n)$ converge to the Law of a r.v. Y.

Condition B: $\varphi_n(\mu_n) > 0$.

Condition C: For an arbitrary closed, bounded interval $[a,b]$,

$$\lim_{n\to\infty} \frac{\varphi_n\left[\dfrac{x}{\varphi_n(\mu_n)} + \mu_n\right]}{\varphi_n(\mu_n)} = 1$$

uniformly for $x \in [a,b]$.

Theorem 1. (Curtiss) Consider the r.v.

$$T_n = \int_{\mu_n}^{X_n} \varphi_n(x)\,dx \quad . \tag{2}$$

If conditions A, B, and C hold, the Laws of T_n converge to the Law of Y.

In the argument below, we only need condition C holding for almost all x. One merely replaces "infimum" and "supremum" by "essential infimum" and "essential supremum".)

(In [3], Curtiss hypothesized and used the continuity of the d.f. $F(\omega)$ of Y.)

Proof of Theorem 1: That T_n is a r.v. is assured by the measurability of a continuous function of a measurable function.

Since $\mathcal{L}(Y_n) \to \mathcal{L}(Y)$, there exist continuity points a and b of F and n_1 such that

(3) $$P(a \leq Y_n \leq b) > 1 - \epsilon$$

For all $n > n_1$. Consider $n > n_1$. From elementary properties of conditional probability

(4) $$P(T_n \leq w | a \leq Y_n \leq b)(1 - \epsilon) \leq P(T_n \leq w) \leq P(T_n \leq w | a \leq Y_n \leq b) + \epsilon \ .$$

(Here $P(\# \mid *)$ is the conditional probability of the event $\#$ given that the event $*$ holds.)

Now the condition $a \leq Y_n \leq b$ is the same as the condition

(5) $$\frac{a}{\varphi_n(\mu_n)} + \mu_n \leq X_n \leq \frac{b}{\varphi_n(\mu_n)} + \mu_n \ .$$

Also

(6) $$m(d - c) \leq \int_c^d f(x)\,dx \leq M(d - c)$$

if $f(x)$ satisfies $m \leq f(x) \leq M$ on the interval $[c,d]$. Thus, for $Y_n \in [a,b]$, so long as $X_n \geq \mu_n$ we may write

(7) $$(X_n - \mu_n)m_n \leq T_n \leq (X_n - \mu_n)M_n$$

where

$$(8) \qquad m_n = \inf_{x\epsilon[a,b]} \varphi_n\left(\frac{x}{\varphi_n(\mu_n)} + \mu_n\right)$$

and

$$(9) \qquad M_n = \sup_{x\epsilon[a,b]} \varphi_n\left(\frac{x}{\varphi_n(\mu_n)} + \mu_n\right) .$$

If $X_n \leq \omega$, M_n and m_n are interchanged in equation 7. Consequently, for $\omega \epsilon [a,b]$,

$$(10) \quad P((X_n - \mu_n) M_n \leq \omega \text{ and } X_n - \mu_n > 0 \mid a \leq Y_n \leq b) + P((X_n - \mu_n) m_n \leq \omega \text{ and } X_n - \mu_n \leq 0 \mid a \leq Y_n \leq b) \leq P(T_n \leq \omega \mid a \leq Y_n \leq b) \leq$$

$$P((X_n - \mu_n) m_n \leq \omega \text{ and } X_n - \mu_n < 0 \mid a \leq Y_n \leq b) + P((X_n - \mu_n) M_n \leq \omega \text{ and } X_n - \mu_n \leq 0 \mid a \leq Y_n \leq b) .$$

But (assuming $a < 0 < b$) this reduces to

$$(11a) \quad \frac{F_n(\omega \frac{\varphi_n(\mu_n)}{M_n}) - F_n(a)}{F_n(b) - F_n(a)} \leq P(T_n \leq \omega \mid a \leq Y_n \leq b) \leq \frac{F_n(\omega \frac{\varphi_n(\mu_n)}{m_n}) - F_n(a)}{F_n(b) - F_n(a)}$$

when ω is positive and to

$$\frac{F_n(\omega \frac{\varphi_n(\mu_n)}{m_n}) - F_n(a)}{F_n(b) - F_n(a)} \leq P(T_n \leq \omega \mid a \leq Y_n \leq b) \leq \frac{F_n(\omega \frac{\varphi_n(\mu_n)}{M_n}) - F_n(a)}{F_n(b) - F_n(a)}$$

when ω is non-positive.

But as an immediate consequence of Condition C we have

$$(12) \qquad \lim_{n \to \infty} \frac{m_n}{\varphi_n(\mu_n)} = 1$$

and

$$(13) \qquad \lim_{n \to \infty} \frac{M_n}{\varphi_n(\mu_n)} = 1 .$$

Thus if ω is a continuity point of F, the first and last terms in (11) can be made arbitrarily close to $F(\omega)$ and the middle term can be made (according to

(4)) arbitrarily close to $P(T_n \leq \omega)$. Thus $\mathcal{L}(T_n) \to \mathcal{L}(Y)$ as asserted.

Actually we can do better. The heuristic argument suggests that the difference $T_n - Y_n$ should be small in some sense. In fact, we can prove

Theorem 2. Under the hypotheses of Theorem 1, the sequence $\{T_n - Y_n\}$ converges to 0 with probability 1 .

Proof: We let ω represent a sample point so that $Z(\omega)$ represents the value of the r.v. Z at the point ω . With the restrictions preceding (7) we have, as a consequence of (7) ,

$$(14) \quad (X_n(\omega)-\mu_n)(m_n - \varphi_n(\mu_n)) = T_n(\omega) - Y_n(\omega) \leq (X_n(\omega)-\mu_n)(M_n - \varphi_n(\mu_n)) \ .$$

(If $X_n(\omega) < \mu_n$, the inequalities are reversed.) In (3) we can always choose a negative and b positive, and we note from (5) that, under the conditions assumed,

$$(15) \qquad \frac{a}{\varphi_n(\mu_n)} \leq X_n(\omega) - \mu_n \leq \frac{b}{\varphi_n(\mu_n)} \ .$$

From (14) and (15) we find that $|T_n(\omega) - Y_n(\omega)|$ is no larger than the largest of the four numbers

$$a\frac{(\varphi_n(\mu_n) - M_n)}{\varphi_n(\mu_n)}, \quad a\frac{(m_n - \varphi_n(\mu_n))}{\varphi_n(\mu_n)}, \quad b\frac{(M_n - \varphi_n(\mu_n))}{\varphi_n(\mu_n)}, \quad b\frac{(\varphi_n(\mu_n) - m_n)}{\varphi_n(\mu_n)}$$

But each of these can be made arbitrarily close to 0 by choosing n sufficiently large. Thus for $Y_n(\omega) \in [a,b]$, $\lim_{n\to\infty} (Y_n(\omega) - T_n(\omega)) = 0$. Thus $P((Y_n - T_n) \to 0) > 1 - \epsilon$ for each positive ϵ and the theorem is proved.

4. Large Neighborhoods and Small. Theorems 1 and 2 require that the functions $\varphi_n(x)$ be nearly constant over arbitrary closed, bounded intervals. They have as corollaries many of the theorems requiring approximate linearity of a related function in a small neighborhood. We prove several such theorems.

Theorem 3. Suppose the sequence of r.v.s $\{X_n - \mu\}$ converges in probability to 0 and $\{(X_n - \mu)\alpha_n\}$ with $\alpha_n > 0$ for each n converges in law to a r.v. Y . Suppose $\xi(x)$, a real-valued function defined on the reals, has a continuous first derivative which does not vanish at $x = \mu$. Then the sequence

$$\left\{\frac{\xi(X_n) - \xi(\mu)}{\xi'(\mu)}\alpha_n\right\}$$

converges in law to Y . (Here and subsequently $\xi'(x)$ is the derivative of $\xi(x)$.

Proof: We suppose the law of Y does not assign measure 1 to a single point for otherwise the theorem is trivial. Since

$$\frac{\alpha_n(\xi(X_n) - \xi(\mu))}{\xi'(\mu)} = \int_{\mu}^{X_n} \frac{\alpha_n \xi'(x)}{\xi'(\mu)}\, dx \tag{16}$$

we need only check that conditions A, B and C are satisfied for the function

$$\varphi_n(x) = \alpha_n \frac{\xi'(x)}{\xi'(\mu)} .$$

Since $\varphi_n(\mu) = \alpha_n$, conditions A and B are satisfied by hypothesis. To check condition C we observe that the sequence α_n diverges to infinity since $\{X_n - \mu\}$ converges in probability to 0 and $(X_n - \mu)\alpha_n$ converges in law to a non-degenerate law. Condition C reduces to

$$\lim_{n\to\infty} \frac{\xi'(\frac{x}{\alpha_n} + \mu)}{\xi'(\mu)} = 1 \tag{17}$$

uniformly for x in $[a,b]$. But since α_n diverges to infinity, this is merely the statement that $\xi'(x)$ is continuous at $x = \mu$, since for n sufficiently large and all $x \in [a,b]$, $\frac{x}{\alpha_n} + \mu$ is in an arbitrary neighborhood of μ.

We only need the continuity of $\xi'(x)$ at μ.

Theorem 3 has as a special case the one-dimensional case of 5e.1 of Rao [8].

The μ of Theorem 3 can be replaced by a sequence of μ_n which converges to a finite limit μ. For we need only to verify that

$$\lim_{n\to\infty} \frac{\xi'(\frac{x}{\alpha_n} \cdot \frac{\xi'(\mu)}{\xi'(\mu_n)} + \mu_n)}{\xi'(\mu_n)} = 1 \tag{18}$$

uniformly for $x \in [a,b]$. Again this is just the continuity of $\xi'(x)$ at μ. Hence we have

Theorem 4. Suppose the sequence of r.v.s $\{X_n - \mu_n\}$ converges in probability to 0, the sequence of real numbers $\{\mu_n\}$ converges to a finite limit μ and $\{(X_n - \mu_n)\alpha_n\}$ with α_n positive converges in law to a r.v. Y.

If $\xi(x)$ has a continuous first derivative which does not vanish at $x = \mu$, than the sequence

$$\left\{ \frac{\xi(X_n) - \xi(\mu_n)}{\xi'(\mu_n)} \alpha_n \right\}$$

converges in law to Y .

Theorem 4 has as a special case the asymptotic normality of "smooth" functions of maximum likelihood estimates when those estimates satisfy the usual regularity conditions guaranteeing asymptotic unbiasedness, consistency and normality. Such a theorem is 12.3.7 of Wilks's book [9].

The question arises as to whether one can use a sequence of functions $F_n(x)$ in Theorem 3. Additional assumptions are necessary.

Theorem 5. Suppose the sequence of r.v.s $\{X_n - \mu\}$ converges in probability to 0 and $\{(X_n - \mu)\alpha_n\}$ converges in law to Y . Suppose $\xi_n(x)$ is a sequence of differentiable functions such that

$$\lim_{n \to \infty} \xi'_n(x) = \eta(x) \tag{19}$$

for each x, with $\eta(\mu) \neq 0$ and such that the functions $\xi'_n(x)$ are equicontinuous, then the sequence

$$\left\{ \frac{\xi_n(X_n) - \xi_n(\mu)}{\eta(\mu)} \alpha_n \right\}$$

converges in law to Y .

Proof: We consider $\varphi_n(x) = \frac{\xi'_n(x)}{\eta(\mu)} \alpha_n$.

Now $\varphi_n(\mu) = \frac{\xi'_n(\mu)}{\eta(\mu)} \alpha_n$, so that Condition A is satisfied since

$$\lim_{n\to\infty} \frac{\xi'_n(\mu)}{\eta(\mu)} = 1$$

and Condition B is satisfied by restricting attention to n large enough that

$$\frac{\xi'_n(\mu)}{\eta(\mu)} > 0 \ .$$

To check Condition C, we need to show that

$$\lim_{n\to\infty} \frac{\xi'_n\left(\frac{x}{\alpha_n} \frac{\eta(\mu)}{\xi'_n(\mu)} + \mu\right)}{\xi'_n(\mu)} = 1 \tag{20}$$

uniformly for $x \in [a,b]$. We show that the numerator can be made arbitrarily close to the non-zero constant $\eta(\mu)$ and that will suffice. We have

$$\left|\xi'_n\left(\frac{x}{\alpha_n} \frac{\eta(\mu)}{\xi'_n(\mu)} + \mu\right) - \eta(\mu)\right| \le \left|\xi'_n\left(\frac{x}{\alpha_n} \frac{\eta(\mu)}{\xi'_n(\mu)} + \mu\right) - \xi'_n(\mu)\right| + \left|\xi'_n(\mu) - \eta(\mu)\right| \tag{21}$$

The first term on the right can be made small by the equicontinuity of $\xi'_n(x)$ at $x = \mu$ and the second can be made small by the convergence of $\xi'_n(\mu)$ to $\eta(\mu)$.

Since we have proved conditions A, B and C we could invoke Theorem 2 rather than Theorem 1, thus showing convergence with probability 1 of the differences.

5. Applications to Poisson Distributions. In this section we consider the problem of normalizing the Poisson distribution -- a problem in which there has been some recent interest [6]. Let X_n have a Poisson distribution with mean n. Then, of course, $\mathcal{L}(\frac{X_n - n}{\sqrt{n}}) \to \mathcal{L}(Y)$ where Y has the standard normal distribution. Further $\frac{X_n}{n}$ converges in probability to the constant 1. A standard theorem [2, §20.6] states that if $\mathcal{L}(Y_n) \to \mathcal{L}(Y)$ and Z_n converges in probability to the non-zero constant c, then $\mathcal{L}(Y_n/Z_n) \to \mathcal{L}(Y/c)$. These three facts can be combined to give the asymptotic normality of a great many r.v.s, e.g.,

$$\sqrt{X_n} - \sqrt{n}\ ,\ \sqrt{X_n + c} - \sqrt{n + c},\ \sum_{n=1}^{k} (\sqrt{X_n + c_n} - \sqrt{n + c_n}),\ \text{and}\ \frac{R(\frac{X_n}{n})}{R(1)} \cdot \frac{(X_n - n)}{\sqrt{n}}\ ,$$

where $R(x)$ is any rational function. The first three are also immediate consequences of Theorem 1. Theorem 1 gives other examples. We give an intuitively appealing form of Curtiss's theorem for r.v.s with Poisson distributions.

Theorem 6. If X_n has a Poisson distribution with mean n and if $\varphi_n(x)$ is a sequence of continuous real-valued functions defined for real x, such that

$$\lim_{n\to\infty} \varphi_n(x\sqrt{n} + n)\sqrt{n} = 1 \tag{22}$$

uniformly on an arbitrary closed and bounded interval $[a,b]$, then the sequence of laws of

$$\int_n^{X_n} \varphi_n(x)\,dx$$

converges to that of a standard normal r.v. Y .

(Roughly this theorem says that a function which is nearly constant within many standard deviations of the mean will serve as integrand for a normalizing transformation.)

To prove the theorem, we show that (22) implies conditions A, B, and C . Taking $x = 0$ in (22) we have

$$\lim_{n\to\infty} \varphi_n(n)\sqrt{n} = 1 \tag{23}$$

so that conditions A and B are satisfied. Thus we need only show that (22) implies condition C . But, in virtue of (23), this reduces to showing that

$$\lim_{n\to\infty} \varphi_n\left(\frac{x}{\varphi_n(n)} + n\right)\sqrt{n} = 1 \tag{24}$$

uniformly for $x \in [a,b]$.

Establishing (24) is an easy application of the "Moore-Osgood" iterated limits theorem with a parameter [4, Theorem VII. 4, p.102]. We consider an arbitrary finite closed interval $[a,b]$ and the double sequence of functions

$$\psi_{m,n}(x) = \varphi_n\left(\frac{x\sqrt{n}}{\varphi_m(m)\sqrt{m}} + \sqrt{n}\right)\sqrt{n} \quad . \tag{25}$$

Since (23) holds, we have, for arbitrary positive ϵ ,

$$\frac{x}{\varphi_m(m)\sqrt{m}} \in [a-\epsilon,\, b+\epsilon]$$

for m sufficiently large, say, $m \geq m_0$. Consequently, by hypothesis,

$$(26) \qquad \lim_{n\to\infty} \psi_{m,n}(x) = 1$$

uniformly for $x \in [a,b]$ and $m \geq m_0$. Since $\varphi_n(x)$ is continuous and hence uniformly continuous on finite closed intervals

$$(27) \qquad \lim_{m\to\infty} \psi_{m,n}(x) = \varphi(x\sqrt{n} + n)\sqrt{n}$$

uniformly for $x \in [a,b]$ for each n. But then by the Moore-Osgood theorem the double limit

$$\lim_{\substack{m\to\infty \\ n\to\infty}} \psi_{m,n}(x)$$

exists uniformly for $x \in [a,b]$ and is equal to the iterated limit

$$\lim_{m\to\infty} \lim_{n\to\infty} \psi_{m,n}(x) .$$

But this iterated limit is 1, according to (26). In particular, then, we can assert that the limit of the "diagonal" terms,

$$(28) \qquad \lim_{n\to\infty} \psi_{n,n}(x) = 1$$

uniformly for $x \in [a,b]$. Now (28) is the same as (24), so the proof is complete.

(If $R(x)$ is any rational function, not vanishing at $x = n$ then $R(x)/(R(n)\sqrt{n})$ satisfies (22). In particular if $R(x) = 1/(x+c)$ one has that

($\sqrt{n}\log(X_n + c) - \sqrt{n}\log(n + c)$) is asymptotically normal. Thus a germ of the reason that $\log X$ is normalizing for Poisson r.v.s is the statement "Out near $n, x/n$ is about 1.")

(There is nothing special about the role of $\sqrt{n}$ or of the Poisson distribution in the above proof. One can prove a general theorem by exactly the same methods.)

6. Final Remarks. The theorems here do not really answer the important questions of data transformation. In fact, they seem to raise more then they solve, for they indicate a tremendous latitude in the choice of the function $\varphi_n(x)$. Thus, if $\varphi_n(x)$ is any function satisfying conditions A, B and C then so does, for instance, $\varphi_n(x)\, R(x)/R(\mu_n)$ where $R(x)$ is any rational function which does not vanish at $x = \mu_n$. But $\varphi_n(x) \equiv \varphi_n(\mu_n)$ being constant satisfies C, so that many transformations will work.

This latitude in the choice of $\varphi_n(x)$, the fact that Theorem 2 says that, asymptotically, one is not changing Y_n, and general remarks about applying asymptotic theorems show that any such transformation should carry with it analysis of the closeness of approximation. The large body of work done on this subject is particularly reassuring.

REFERENCES

[1] Arley, Niels and Buch, K. Rander (1950). Introduction to the Theory of Probability and Statistics. John Wiley and Sons, Inc., New York.

[2] Cramér, Harald (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton.

[3] Curtiss, J. H. (1943). On transformations used in the analysis of variance. Ann. Math. Statist. **14**. 107-122.

[4] Graves, Lawrence M. (1956). The Theory of Functions of a Real Variable. (2nd Ed.). McGraw-Hill Book Co., Inc., New York.

[5] Keeping, E. S. (1962). Introduction to Statistical Inference. D. van Nostrand Co., Inc., Princeton.

[6] Laubscher, Nico F. (1960). On the stabilization of Poisson variance. Trabajos Estadist. **11**. 199-207.

[7] Parzen, Emanuel (1960). Modern Probability Theory and Its Applications. John Wiley and Sons, Inc., New York.

[8] Rao, C. Radhakrishna (1952). Advanced Statistical Methods In Biometric Research. John Wiley and Sons, Inc., New York.

[9] Wilks, Samuel S. (1962). Mathematical Statistics. John Wiley and Sons, Inc., New York.